ELSEVIER

Chemometrics and Intelligent Laboratory Systems 47 (1999) 1–16

Chemometrics and intelligent laboratory systems

Tutorial

Rough sets theory

B. Walczak [*,1], D.L. Massart

ChemoAC, Pharmaceutical Institute, Vrije Universiteit Brussel, Laarbeeklaan 103, B-1090 Brussels, Belgium

Accepted 1 December 1998

Abstract

The basic concepts of the rough set theory are introduced and adequately illustrated. An example of the rough set theory application to the QSAR classification problem is presented. Numerous earlier applications of rough set theory to the various scientific domains suggest that it also can be a useful tool for the analysis of inexact, uncertain, or vague chemical data.

Keywords: Set theory; Classification; QSAR; Rough sets; Supervised pattern recognition

Contents

* Corresponding author. Silesian University, 9 Szkolna Street, 40-006 Katowice, Poland
[1] On leave from the Silesian University, 40-006 Katowice, 9 Szkolna Street, Poland.

0169-7439/99/$ - see front matter
PII: S0169-7439(98)00200-7

1. Introduction

Often, information on the surrounding world is imprecise, incomplete or uncertain. Still our way of thinking and concluding depends on information at our disposal. This means that to draw conclusions, we should be able to process uncertain and/or incomplete information.

Tools, which turned out to be particularly adequate for the analysis of various types of data, especially, when dealing with inexact, uncertain or vague knowledge, are the fuzzy set and the rough set theories. Rough sets and fuzzy sets capture two distinct aspects of imperfection in knowledge: indiscernibility (a term, which we will define later) and vagueness. The fuzzy set theory, introduced by Zadeh in 1965 [1], has already demonstrated its usefulness in chemistry and in other disciplines [2–9]. The rough set theory, introduced by Pawlak in 1985 [10,11], although popular in many other disciplines [12], is nearly unknown in chemistry [13,14]. What is unique about this approach when compared with the other approaches to the set theory?

In the ordinary set theory, crisp sets are used. A set is then defined uniquely by its elements, i.e., to define a set we have to point out its elements. The membership function, describing the belongingness of elements of the universe to the set, can attain one of the two values, 0 or 1 (see Fig. 1a).

It means that any element is either in or outside the set under consideration. This definition of the membership function does not take into the account the uncertainty of being an element of a given set of elements.

To deal with the uncertainty problems, the concept of fuzzy set was introduced. Fuzzy set is defined by the membership function which can attain values from the closed interval [0,1], allowing partial membership of the elements in the set (see Fig. 1b). Fuzziness measures the degree to which an event occurs and not whether it occurs.

In the rough set theory, membership is not the primary concept. Rough sets represent a different mathematical approach to vagueness and uncertainty. Definition of a set in the rough set theory is related to our information (knowledge) and perception about elements of the universe. In other words, we 'see' elements of the universe in the context of an available information about them. As a consequence, two different elements can be indiscernible in the context of the information about them and 'seen' as the same. Consider a simple example. Two acids with pKs of respectively pK 4.12 and 4.53 will, in many contexts, be perceived as so equally weak, that they are indiscernible with respect to this attribute. They are part of a rough set 'weak acids' as compared to 'strong' or 'medium' or whatever other category, relevant to the context of this classification.

The rough set methodology is based on the premise that lowering the degree of precision in the data makes the data pattern more visible [15], whereas the central premise of the rough set philosophy is that the knowledge consists in the ability of classification. In other words, the rough set approach can be considered as a formal framework for discovering facts from imperfect data. The results of the rough set approach are presented in the form of classification or decision rules derived from a set of examples. For this

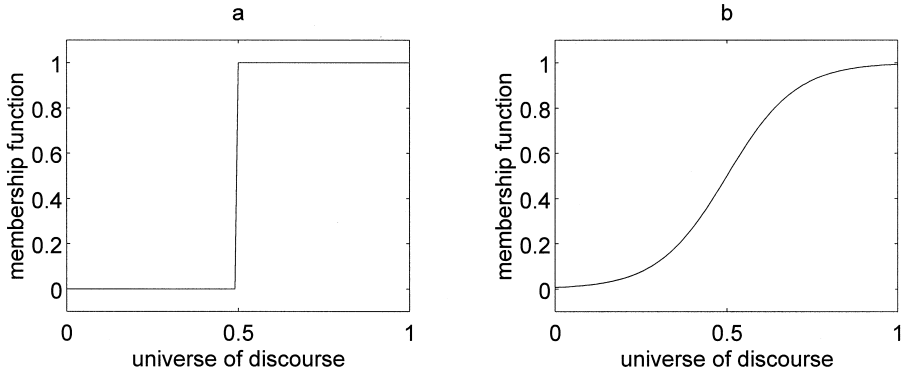

Fig. 1. Membership function according to (a) the crisp set theory and (b) the fuzzy set theory.

reason, and although the approach is very different, it has some aspects in common with inductive rule building methods (e.g., the Id3 algorithm [16]), which are better known to chemists.

The aim of this tutorial is to introduce the basic concepts of the Rough Set Theory (RST) and also to show its possible applications. The terminology of the RST is very specific, but as the readers who are interested in further readings will need to understand it, we decided to use it in our tutorial.

2. Theory

2.1. Basic concepts of the rough sets theory

2.1.1. Information system

Formally, an *information system*, IS (or an approximation space), can be seen as a system

$$\mathrm{IS} = (U, A)$$

where U is the universe (a finite set of objects, $U = \{x_1, x_2, \ldots, x_m\}$) and A is the set of attributes (features, variables). Each attribute $a \in A$ (attribute a belonging to the considered set of attributes A) defines an information function f_a: $U \rightarrow V_a$, where V_a is the set of values of a, called the domain of attribute a.

Example 1

Consider a data set containing the results of three measurements performed for 10 objects. The results can be organized in a matrix (10×3).

2	1	3
3	2	1
2	1	3
2	2	3
1	1	4
1	1	2
3	2	1
1	1	4
2	1	3
3	2	1

Using the terminology of the rough sets theory, this data set can be considered as an information system IS = (U, A), where universe U and attributes A correspond to the set of objects and to the set of variables, respectively:

$$U = \{x_1, x_2, x_3, x_4, x_5, x_6, \ldots, x_{10}\}$$

$$A = \{a_1, a_2, a_3\}.$$

The domains of the particular attributes are:

$$V_1 = \{1,2,3\}$$

$$V_2 = \{1,2\}$$

$$V_3 = \{1,2,3,4\},$$

i.e., the domain of each attribute is the set of values of this attribute. The information function f_a for this system is presented in Table 1.

2.1.2. Indiscernibility relation

For every set of attributes $B \subset A$, an *indiscernibility relation* Ind(B) is defined in the following way: two objects, x_i and x_j, are indiscernible by the set of attributes B in A, if $b(x_i) = b(x_j)$ for every $b \subset B$. The equivalence class of Ind(B) is called *elementary set* in B because it represents the smallest discernible groups of objects. For any element x_i of U, the equivalence class of x_i in relation Ind(B) is represented as $[x_i]_{\mathrm{Ind}(B)}$. The construction of elementary sets is the first step in classification with rough sets.

Example 2

As one can easily notice, there are some identical objects in our data set. For instance, objects x_1 and x_3 cannot be distinguished based on the available data.

Let us group all objects based on the three variables considered. The results are presented in Table 2.

Table 1

U	a_1	a_2	a_3
x_1	2	1	3
x_2	3	2	1
x_3	2	1	3
x_4	2	2	3
x_5	1	1	4
x_6	1	1	2
x_7	3	2	1
x_8	1	1	4
x_9	2	1	3
x_{10}	3	2	1

Each row in this table describes one elementary set, whereas the whole table describes the IS studied. The notation U/A means that we are considering elementary sets of the universe U in the space A.

It can happen that we are interested in the two attributes only, for instance in a_1 and a_2. Then the indiscernibility relation is limited to the subset $B = \{a_1, a_2\}$ and the resulting elementary sets are given in Table 3.

Table 3

U/B	a_1	a_2
$\{x_1, x_3, x_9\}$	2	1
$\{x_2, x_7, x_{10}\}$	3	2
$\{x_4\}$	2	2
$\{x_5, x_6, x_8\}$	1	1

2.1.3. Lower and upper approximations

The rough sets approach to data analysis hinges on two basic concepts, namely the *lower* and the *upper approximations* of a set (Fig. 2), referring to:

-the elements that doubtlessly belong to the set, and

-the elements that possibly belong to the set.

Let X denote the subset of elements of the universe U ($X \subset U$). The lower approximation of X in B ($B \subseteq A$), denoted as $\underline{\mathrm{BX}}$, is defined as the union of all these elementary sets which are contained in X. More formally:

$$\underline{\mathrm{BX}} = \left\{ x_i \in U | \left[x_i \right]_{\mathrm{Ind}(B)} \subset X \right\}.$$

The above statement is to be read as: the lower approximation of the set X is a set of objects x_i, which belong to the elementary sets contained in X (in the space B).

The upper approximation of the set X, denoted as BX, is the union of these elementary sets, which have a non-empty intersection with X:

$$\mathrm{BX} = \left\{ x_i \in U | \left[x_i \right]_{\mathrm{Ind}(B)} \cap X \neq 0 \right\}.$$

For any object x_i of the lower approximation of X (i.e., $x_i \in \underline{\mathrm{BX}}$), it is certain that it belongs to X. For any object x_i of the upper approximation of X (i.e., $x_i \in \mathrm{BX}$), we can only say that x_i may belong to X. The difference:

$$\mathrm{BNX} = \mathrm{BX} - \underline{\mathrm{BX}}$$

is called a boundary of X in U.

Table 2

U/A	a_1	a_2	a_3
$\{x_1, x_3, x_9\}$	2	1	3
$\{x_2, x_7, x_{10}\}$	3	2	1
$\{x_4\}$	2	2	3
$\{x_5, x_8\}$	1	1	4
$\{x_6\}$	1	1	2

If the lower and upper approximation are identical (i.e., $\mathrm{BX} = \underline{\mathrm{BX}}$), then set X is definable, otherwise, set X is undefinable in U. There are four types of undefinable sets in U:

1. if $\underline{\mathrm{BX}} \neq \emptyset$ and $\mathrm{BX} \neq U$, X is called roughly definable in U;
2. if $\underline{\mathrm{BX}} \neq \emptyset$ and $\mathrm{BX} = U$, X is called externally undefinable in U;
3. if $\underline{\mathrm{BX}} = \emptyset$ and $\mathrm{BX} \neq U$, X is called internally undefinable in U;
4. if $\underline{\mathrm{BX}} = \emptyset$ and $\mathrm{BX} = U$, X is called totally undefinable in U,

where $\emptyset$ denotes an empty set.

Additionally, the following notation can be introduced: $\mathrm{POS}_B(X) = \underline{\mathrm{BX}}$, called the B-positive region of X, is the set of these objects, which can, with certainty, be classified in the set X, $\mathrm{NEG}_B(X) = U - \underline{\mathrm{BX}}$, called the B-negative region of X, is the set of objects, which without ambiguity, can be classified as belonging to the complement of X (or as not belonging to X), $\mathrm{BN}_B(X)$, called the B-borderline region of X, is an undecidable area of the universe, i.e., none of the objects belonging to the boundary can, with certainty, be classified into X or $-X$, as far as the attributes B are considered.

Example 3

Let us assume that we are interested in the subset X of five objects $\{X = x_1, x_3, x_4, x_5, x_9\}$. Can we distinguish this set from the whole data set in the space of three attributes ($B = \{a_1, a_2, a_3\}$)? Based on the results presented in Table 2, one can calculate the lower and upper approximations of this set in the following way.

The elementary sets presented in Table 2, which are also contained in X, are:

$$\left\{ x_1, x_3, x_9 \right\}, \ \left\{ x_4 \right\}.$$

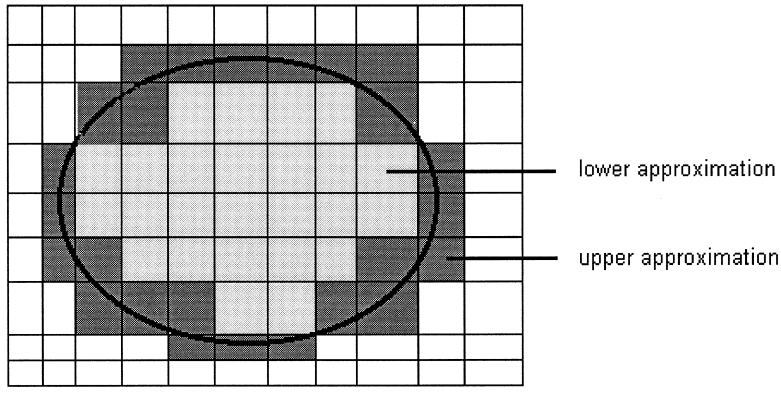

Fig. 2. Schematic demonstration of the upper and lower approximation of set X.

It means that the lower approximation is given by the following set of objects:

$$\underline{\mathrm{BX}} = \{x_1, x_3, x_4, x_9\}.$$

To calculate the upper approximation of the subset X, one has to find in Table 2 all elementary sets which have at least 1 element in common with the subset X. These are:

$$\{x_1, x_3, x_9\},\ \{x_4\},\ \{x_5, x_8\}$$

so that the upper approximation is:

$$\mathrm{BX} = \{x_1, x_3, x_4, x_5, x_8, x_9\}.$$

The boundary of X in U, defined as the difference between the upper and lower approximations, contains elements which are in the upper but not in the lower approximation:

$$\mathrm{BNX} = \{x_1, x_3, x_4, x_5, x_8, x_9\} - \{x_1, x_3, x_4, x_9\} = \{x_5, x_8\}.$$

2.1.4. Accuracy of approximation

An accuracy measure of the set X in $B \subseteq A$ is defined as:

$$\mu_B(X) = \mathrm{card}(\underline{\mathrm{BX}})/\mathrm{card}(\mathrm{BX})$$

The cardinality of a set is the number of objects contained in the lower (upper) approximation of the set X. As one can notice, $0 \le \mu_B(X) \le 1$. If X is definable in U then $\mu_B(X) = 1$, if X is undefinable in U then $\mu_B(X) < 1$.

Example 4

The number of objects contained in the lower approximation of example 3 equals 4. The cardinality of the upper approximation equals 6. The accuracy of set X therefore is: $\mu_B(X) = 4/6$.

This means that the considered set X is roughly definable in U; it can be defined by its lower and upper approximations in U.

2.1.5. Independence of attributes

In order to check, whether the set of attributes is independent or not, one checks for every attribute whether its removal increases the number of elementary sets in the IS or not.

If $\mathrm{Ind}(A) = \mathrm{Ind}(A - a_i)$, then the attribute a_i is called superfluous. Otherwise, the attribute a_i is indispensable in A.

Example 5

Consider Table 1. If the three attributes (a_1, a_2 and a_3) are taken into account, five elementary sets can be constructed (see Table 2). Table 4 gives the number of elementary sets after leaving out one of the attributes. For instance, if only a_2 and a_3 are used, five elementary sets are distinguished, if a_1 and a_3 are used, the number of elementary sets is 4.

If we remove attribute a_2 or a_3, the number of elementary sets becomes smaller, but by removing attribute a_1, we do not change the elementary sets. Attribute a_1 is superfluous, whereas attributes a_2 and a_3 are indispensable.

Table 4

	Removed attribute			
	None	a_1	a_2	a_3
Number of elementary sets	5	5	4	4

The set of attributes is dependent because by removing attribute a_1, we obtain the information system identical with that presented in Table 2.

Elimination of superfluous attributes simplifies the information set and has diagnostic value. It should be noted here that later, we will define so-called *D*-superfluous attributes. This definition is used for classification purpose.

2.1.6. Core and reduct of attributes

If the set of attributes is dependent, one can be interested in finding all possible minimal subsets of attributes, which lead to the same number of elementary sets as the whole set of attributes (*reducts*) and in finding the set of all indispensable attributes (*core*).

The concepts of core and reduct are two fundamental concepts of the rough sets theory. The reduct is the essential part of an IS, which can discern all objects discernible by the original IS. The core is the common part of all reducts. To compute reducts and core, the *discernibility matrix* is used. The discernibility matrix has the dimension $n \times n$, where n denotes the number of elementary sets and its elements are defined as the set of all attributes which discern elementary sets $[x]_i$ and $[x]_j$.

Example 6

The discernibility matrix, D, for the five elementary sets presented in Table 2, is constructed in the following way. To calculate element d_{ij}, one ought to find the set of attributes which discern the elementary sets i and j. The set of attributes which discern the elementary sets 1 and 2 contains attributes a_1, a_2, and a_3, i.e., $d_{21} = d_{12} = \{a_1, a_2, a_3\}$. The element $d_{31} = d_{13} = \{a_2\}$, i.e., the attribute a_2 only discerns the elementary sets 3 and 1. As the discernibility matrix is symetrical ($d_{ij} = d_{ji}$), it is enough to consider its lower diagonal part only.

Of course, each elementary set differs from the rest of elementary sets, due to at least one attribute, so that there are no empty cells in the discernibility matrix.

The discernibility matrix can be used to find the minimal subset(s) of attributes (reduct(s)), which leads to the same partition of the data as the whole set of attributes A. To do this, one has to construct the so-called discernibility function $f(A)$. This is a Boolean function, constructed in the following way: to each attribute from the set of attributes, which discern two elementary sets, (e.g., $\{a_1, a_2, a_3\}$), we assign a Boolean variable 'a', and the resulting Boolean function attains the form $(a_1 + a_2 + a_3)$ (or it can be presented as $(a_1 \vee a_2 \vee a_3)$). If the set of attributes is empty, we assign to it the Boolean constant 1. For the discernibility matrix presented in Table 5, the discernibility function has the following form:

$$f(A) = (a_1 + a_2 + a_3)\, a_2 (a_1 + a_3)(a_1 + a_3) \times (a_1 + a_3)(a_1 + a_2 + a_3)(a_1 + a_2 + a_3) \times (a_1 + a_2 + a_3)(a_1 + a_2 + a_3)\, a_3.$$

To calculate the final form of $f(A)$, the absorption law is applied. According to the absorption law, if the elementary set 1 differs from the elementary set 2 due to the attributes a_1, a_2 and a_3, and from the elementary set 3 due to the attribute a_2, it is enough to take into the account the attribute a_2 only, which discerns this set from both set 2 and set 3, i.e:

$$(a_1 + a_2 + a_3)\, a_2 = a_2$$

Let us look at another example. Suppose that to discern the elementary set 1 from the sets 2, 3, 4 and 5, one has to take into account the following sets of attributes (see the first column in Table 5):

$$\{a_2\},\ \{a_1, a_2, a_3\},\ \{a_1, a_3\} \text{ and } \{a_1, a_3\}$$

Then, the discernibility function attains the following form:

$$a_2(a_1 + a_2 + a_3)(a_1 + a_3)(a_1 + a_3) = a_2(a_1 + a_3) = a_2 a_1 + a_2 a_3$$

The right hand side of the above equation represents what is called 'disjunctive normal form of $f(A)$'. A Boolean expression is said to be in the normal form, if it is composed of Boolean variables and constants only, linked by operators of disjunction ($\vee$)

Table 5

	Set 1	Set 2	Set 3	Set 4	Set 5
Set 1					
Set 2	a_1, a_2, a_3				
Set 3	a_2	a_1, a_3			
Set 4	a_1, a_3	a_1, a_2, a_3	a_1, a_2, a_3		
Set 5	a_1, a_3	a_1, a_2, a_3	a_1, a_2, a_3	a_3	

Table 6

U/R	a_2	a_3
$\{x_1, x_3, x_9\}$	1	3
$\{x_2, x_7, x_{10}\}$	2	1
$\{x_4\}$	2	3
$\{x_5, x_8\}$	1	4
$\{x_6\}$	1	2

and conjunction ($\wedge$). A Boolean expression is in conjunctive normal form, if it is a conjunction of a number of disjuncts, for instance:

$$(a_1 + a_2 + a_3)(a_1 + a_3)$$

A Boolean expression is in disjunctive normal form, if it is a disjunction of a number of constituents, each of them being a conjunction, for instance:

$$a_1 a_2 + a_1 a_3$$

The indiscernibility function in its disjunctive normal form is an alternative representation of the IS.

In practice, calculation of core and reduct(s) is very simple, as core is the set of all single element entries of the discernibility matrix, whereas reduct B' is the minimal subset of attributes such that it has at least one common element with any non-empty entry in the discernibility matrix.

Let us return to our example. For the discernibility matrix (see Table 5), we obtained:

$$f(A) = a_2 a_3$$

In this way, we found one reduct of the considered set of attributes. Usually, there are a few reducts, which allow to represent the IS in alternative ways. As we found only one reduct, we can represent our system as in Table 6.

This representation is equivalent to that in Table 2. Attributes a_2 and a_3 provide the same knowledge (partition) about the universe U, as the sets of attributes a_1, a_2 and a_3.

This example demonstrates that the aim of finding reducts is to find alternative presentations of the information system. It also means that the reduct, R, which is the minimal subset of independent attributes (in our example $R = \{a_2, a_3\}$), leads to the same partition of the data as the whole set of attributes A (here $A = \{a_1, a_2, a_3\}$), i.e., $\mathrm{Ind}(R) = \mathrm{Ind}(A)$.

The core, as defined earlier, is the set of all single attributes occurring in the discernibility matrix. The core is the set of attributes a_2 and a_3, which means that in this particular case the reduct equals the core.

At this stage, the reader may have some difficulties in distinguishing between eliminating superfluous attributes and determining reduct and core, as they all yield the same results. This is no longer the case in the QSAR-application, which we give at the end of the tutorial.

2.1.7. Core and reducts of attribute values

Simplification of the IS can be achieved by dropping certain values of attributes, which are unnecessary for the system, i.e., by eliminating some of these *values* in such a way that we are still able to discern all elementary sets in the system. The procedure of finding core and reducts of the attribute values is similar to that of finding core and reducts of the attributes. All computations are performed based on the discernibility matrix, but the definition of the discernibility function is now slightly different. Instead of one discernibility function, we have to construct as many discernibility functions, as there are elementary sets in the IS.

Example 7

Based on Table 6, we can construct five discernibility functions $f_1(A), f_2(A), \ldots, f_5(A)$ for five elementary sets in the space of attributes $\{a_2, a_3\}$. The function $f_1(A)$ concerns the sets of attributes which discern the elementary set 1 from the sets 2, 3, 4 and 5; the function $f_2(A)$, those which discern the elementary set 2 from the sets 1, 3, 4 and 5, etc. For the reader's convenience, the discernibility matrix of Table 5 is presented in a form allowing fast construction of the discussed functions.

To construct $f_1(A)$, one takes into account all sets of attributes from column 1 in Table 7; to construct

Table 7

	Set 1	Set 2	Set 3	Set 4	Set 5
Set 1		a_2, a_3	a_2	a_3	a_3
Set 2	a_2, a_3		a_3	a_2, a_3	a_2, a_3
Set 3	a_2	a_3		a_2, a_3	a_2, a_3
Set 4	a_3	a_2, a_3	a_2, a_3		a_3
Set 5	a_3	a_2, a_3	a_2, a_3	a_3	

Table 8

U/R	a_2	a_3
$\{x_1, x_3, x_9\}$	1	3
$\{x_2, x_7, x_{10}\}$	*	1
$\{x_4\}$	2	3
$\{x_5, x_8\}$	*	4
$\{x_6\}$	*	2

* Denotes 'do not care'.

$f_2(A)$, one takes into account all sets of attributes from column 2, etc.

$$f_1(A) = (a_2 + a_3)\, a_2\, a_3\, a_3 = a_2\, a_3$$

$$f_2(A) = (a_2 + a_3)\, a_3 (a_2 + a_3)(a_2 + a_3) = a_3$$

$$f_3(A) = a_2\, a_3 (a_2 + a_3)(a_2 + a_3) = a_2\, a_3$$

$$f_4(A) = a_3 (a_2 + a_3)(a_2 + a_3)\, a_3 = a_3$$

$$f_5(A) = a_3 (a_2 + a_3)(a_2 + a_3)(a_2 + a_3)\, a_3 = a_3$$

$f_1(A)$ allows to conclude that there is one reduct of the value of attributes, namely $a_2 a_3$. This reduct suggests that we need to care about the values of attribute a_2 and a_3. $f_2(A)$ suggests that we do not need to care for the values of attribute a_2. Consequently, one can simplify Table 6 in the following way (Table 8).

2.1.8. Classification

Let $F = \{X_1, X_2, \ldots, X_n\}$, $X_i \subset U$, be a family of subsets of the universe U. If the subsets in F do not overlap, i.e.,

$$X_i \cap X_j = \emptyset,$$

and the entity of them contains all elementary sets, i.e.,

$$\cup X_i = U \quad \text{for } i = 1, \ldots, n$$

then, F is called a classification of U, whereas X_i are called classes.

The lower and upper approximations of F in $B \subseteq A$ are defined as:

$$\underline{B}(F) = \{\underline{B}(X_1), \underline{B}(X_2), \ldots\ldots, \underline{B}(X_n)\}$$

$$B(F) = \{B(x_1), B(X_2), \ldots\ldots, B(X_n)\},$$

respectively. The quality of classification is defined as:

$$\eta_B F = \cup \mathrm{card}\underline{B}(X_i))/\mathrm{card}U$$

and the accuracy of classification F in B can be calculated according to the following formula:

$$\beta_B F = \cup \mathrm{card}\underline{B}(X_i)/\cup \mathrm{card}\, B(X_i)$$

Example 8

Let us assume that we intend to classify the following three sets of objects using the data of Table 1:

$X_1 = \{x_1, x_3, x_9\}$, $X_2 = \{x_4, x_5, x_6\}$, and

$X_3 = \{x_2, x_7, x_8, x_{10}\}$.

Lower and upper approximations of each class, and the accuracy of their classification are presented in Table 9.

The last column in Table 9 contains accuracy calculated for each class separately. For instance, accuracy of class number 2 is 0.5, because there are two objects in its lower approximation and four objects in its upper approximation ($2/4 = 0.5$). As one can see, class 1 is describable in the system, whereas classes 2 and 3 are roughly describable with accuracies of 0.5 and 0.6, respectively. The accuracy of the whole classification is:

$$\eta_B F = \cup \mathrm{card}\,\underline{B}(X_i)/\cup \mathrm{card}\, B(X_i)$$

$$= (3 + 2 + 3)/(3 + 4 + 5) = 0.67$$

and the quality of the whole classification is:

$$\eta_B F = \cup \mathrm{card}\,\underline{B}(X_i))/\mathrm{card}U$$

$$= (3 + 2 + 3)/(3 + 3 + 4) = 0.8.$$

One can also calculate the accuracy of classification for each class, when removing one attribute from the reduct (see Table 10).

The results shown in Table 10 allow to estimate how the attributes influence accuracy of class description. Removal of attribute a_2 from the reduct causes dramatical changes in the accuracy of classes 1 and 2, but the accuracy of class 3 does not change.

Table 9

Class number	Number of objects	Lower approximation	Upper approximation	Accuracy
1	3	3	3	1.0
2	3	2	4	0.5
3	4	3	5	0.6

Table 10

Class number	Removed attribute		
	None	a_2	a_3
1	1	internally undefinable	internally undefinable
2	0.5	0.14	totally undefinable
3	0.6	0.60	totally undefinable

When attribute a_3 is removed, all the classes become undefinable (class 1 becomes internally undefinable, whereas classes 2 and 3 become totally undefinable).

2.2. Decision table

A knowledge representation system containing the set of attributes A (now called condition attributes) and the set of decision attributes D is called a decision table. As we will show further, decision tables are also useful for classification.

Example 9

The decision table (Table 11) contains three condition attributes $\{a_1, a_2, a_3\}$ and one decision attribute d. The decision attribute d describes the belongingness of 10 objects to the three following classes:

class 1: $\{x_1, x_3, x_9\}$,
class 2: $\{x_2, x_4, x_7, x_{10}\}$,
class 3: $\{x_5, x_6, x_8\}$.

This example demonstrates that any supervised classification problem can be considered as the decision table analysis. However, the concept of decision table analysis is more general than that of data classification. The decision table can contain many decision attributes. For instance, in Table 12 there are three condition attributes $A = \{a_1, a_2, a_3\}$, and two decision attributes $D = \{d_1, d_2\}$.

Table 11

U	a_1	a_2	a_3	d
x_1	2	1	3	1
x_2	3	2	1	2
x_3	2	1	3	1
x_4	2	2	3	2
x_5	1	1	4	3
x_6	1	1	2	3
x_7	3	2	1	2
x_8	1	1	4	3
x_9	2	1	3	1
x_{10}	3	2	1	2

Table 12

U	a_1	a_2	a_3	d_1	d_2
x_1	2	1	3	2	3
x_2	3	2	1	3	1
x_3	2	1	3	2	3
x_4	2	2	3	3	1
x_5	1	1	4	1	3
x_6	1	1	2	1	3
x_7	3	2	1	3	1
x_8	1	1	4	1	3
x_9	2	1	3	2	3
x_{10}	3	2	1	3	1

This type of decision table can be used to study if the properties of objects expressed in terms of attributes A can be expressed in terms of attributes D. This allows, for instance, to compare two knowledge bases or to express knowledge basis D in terms of knowledge base A.

Analysis of the decision table requires the introduction of several new definitions.

2.2.1. D-superfluous attributes

Attribute a_i, belonging to the condition set of attributes B (where $B \subseteq A$), is D-superfluous if it exerts no influence on the lower approximation of D, i.e., if

$$\mathrm{POS}_B(D) = \mathrm{POS}_{(B-a_i)}(D)$$

Otherwise, attribute a_i is D-indispensable in A.

2.2.2. Relative core and relative reducts of attributes

The set of all D-indispensable attributes in A is called the D-core of A, whereas, the minimal subsets of condition attributes that discern all equivalence classes of the relation $\mathrm{Ind}(D)$ discernable by the entire set of attributes are called D-reducts.

Relative reducts can be computed using a slightly modified discernibility matrix. An element of the D-discernibility matrix of A is defined as the set of all attributes which discern the objects x_i and x_j, which do not belong to the same equivalence class of the relation $\mathrm{Ind}(D)$, i.e., to the same class. The D-core is the set of all single elements of the D-discernibility matrix of A.

2.3. Main steps of decision table analysis

1. Construction of elementary sets in D-space,
2. calculation of upper and lower approximations of the elementary sets in D,
3. finding D-core and D-reducts of A attributes,
4. finding D-core and D-reducts of A attribute values.

Example 10

Let us start with Table 12. In the D-space, one can find the following elementary sets:

set 1: $\{x_1, x_3, x_9\}$,
set 2: $\{x_2, x_4, x_7, x_{10}\}$,
set 3: $\{x_5, x_6, x_8\}$.

The two conditions attributes, d_1 and d_2, can be temporarily replaced by a new condition attribute d describing elementary sets in D.

Decision attribute d has a value of 1 for all objects belonging to the elementary set 1, value of 2 for the objects belonging to the elementary sets 2, etc. As one can notice, Table 13 is the same as Table 11, representing the classification problem. It means that all steps of further analysis are the same for both decision tables (Tables 11 and 12).

The results presented in Table 14 indicate that all classes (elementary sets in D) can be properly described in terms of the condition attributes $A = \{a_1, a_2, a_3\}$.

The accuracy of the whole classification and its quality are equal to 1.0.

To find D-core and D-reduct of attributes A, one must first construct the D-discernibility matrix, elements of which discern objects from different groups in D.

Table 13

U	a_1	a_2	a_3	d
x_1	2	1	3	1
x_2	3	2	1	2
x_3	2	1	3	1
x_4	2	2	3	2
x_5	1	1	4	3
x_6	1	1	2	3
x_7	3	2	1	2
x_8	1	1	4	3
x_9	2	1	3	1
x_{10}	3	2	1	2

Table 14

Class number	Number of objects	Lower approximation	Upper approximation	Accuracy
1	3	3	3	1.0
2	4	4	4	1.0
3	3	3	3	1.0

As object x_1 appears in the same class as objects x_3 and x_9 in the D-space, we are not interested in the set of attributes which discern these three objects. For this reason, in the first column of Table 15, rows 1, 3 and 9 are not considered.

The $f_A(D)$ discernibility function has the following form:

$$\begin{aligned} f_A(D) = {} & (a_1 + a_2 + a_3)\, a_2 (a_1 + a_3)(a_1 + a_3) \\ & \times (a_1 + a_2 + a_3)(a_1 + a_3)(a_1 + a_2 + a_3) \\ & \times (a_1 + a_2 + a_3)(a_1 + a_2 + a_3) \\ & \times (a_1 + a_2 + a_3)(a_1 + a_2 + a_3) \\ & \times (a_1 + a_2 + a_3)\, a_2 (a_1 + a_3)(a_1 + a_3) \\ & \times (a_1 + a_2 + a_3)(a_1 + a_3)(a_1 + a_2 + a_3) \\ & \times (a_1 + a_2 + a_3)(a_1 + a_2 + a_3) \\ & \times (a_1 + a_2 + a_3)\, a_2 (a_1 + a_3) \\ & \times (a_1 + a_2 + a_3)(a_1 + a_3)(a_1 + a_2 + a_3) \\ & \times (a_1 + a_2 + a_3)(a_1 + a_3)(a_1 + a_2 + a_3) \\ & \times (a_1 + a_2 + a_3)(a_1 + a_2 + a_3)(a_1 + a_3) \\ & \times (a_1 + a_2 + a_3)(a_1 + a_2 + a_3) \\ = {} & a_2 (a_1 + a_3) = a_1 a_2 + a_2 a_3. \end{aligned}$$

There are two D-reducts $\{a_1, a_2\}$ and $\{a_2, a_3\}$, and the D-core equals $\{a_2\}$. It means that the decision Table 12 can be reduced and presented in two alternative ways as Tables 16 and 17.

We can also be interested in elimination of unnecessary values of condition attributes in the decision table. To do this, we need to calculate a relative reduct and the relative core of values of attributes, based on the D-discernibility matrix constructed for subspace $\{a_1, a_2\}$ or $\{a_2, a_3\}$. In Table 18, the D-discernibility matrix for reduct $\{a_1, a_2\}$ is presented.

Table 15

	1	2	3	4	5	6	7	8	9	10
1	–									
2	a_1,a_2,a_3	–								
3	–	a_1,a_2,a_3	–							
4	a_2	–	a_2	–						
5	a_1,a_3	a_1,a_2,a_3	a_1,a_3	a_1,a_2,a_3	–					
6	a_1,a_3	a_1,a_2,a_3	a_1,a_3	a_1,a_2,a_3	–	–				
7	a_1,a_2,a_3	–	a_1,a_2,a_3	–	a_1,a_2,a_3	a_1,a_2,a_3	–			
8	a_1,a_3	a_1,a_2,a_3	a_1,a_3	a_1,a_2,a_3	–	–	a_1,a_2,a_3	–		
9	–	a_1,a_2,a_3	–	a_2	a_1,a_3	a_1,a_3	a_1,a_2,a_3	a_1,a_3	–	
10	a_1,a_2,a_3	–	a_1,a_2,a_3	–	a_1,a_2,a_3	a_1,a_2,a_3	–	a_1,a_2,a_3	a_1,a_2,a_3	–

The relative discernibility functions are:

$$f_1(D) = (a_1 + a_2)\,a_2\,a_1\,a_1(a_1 + a_2)\,a_1(a_1 + a_2) = a_1 a_2$$

$$f_2(D) = (a_1 + a_2)(a_1 + a_2)(a_1 + a_2)(a_1 + a_2) \times (a_1 + a_2)(a_1 + a_2) = a_1 + a_2$$

$$f_3(D) = (a_1 + a_2)\,a_2\,a_1\,a_1(a_1 + a_2)\,a_1(a_1 + a_2) = a_1 a_2$$

$$f_4(D) = a_2\,a_2(a_1 + a_2)(a_1 + a_2)(a_1 + a_2)\,a_2 = a_2$$

$$f_5(D) = a_1(a_1 + a_2)\,a_1(a_1 + a_2)(a_1 + a_2) \times a_1(a_1 + a_2) = a_1$$

$$f_6(D) = a_1(a_1 + a_2)\,a_1(a_1 + a_2)(a_1 + a_2) \times a_1(a_1 + a_2) = a_1$$

$$f_7(D) = (a_1 + a_2)(a_1 + a_2)(a_1 + a_2)(a_1 + a_2) \times (a_1 + a_2)(a_1 + a_2) = a_1 + a_2$$

$$f_8(D) = a_1(a_1 + a_2)\,a_1(a_1 + a_2)(a_1 + a_2) \times a_1(a_1 + a_2) = a_1$$

$$f_9(D) = (a_1 + a_2)\,a_2\,a_1\,a_1(a_1 + a_2)\,a_1(a_1 + a_2) = a_1 a_2$$

$$f_{10}(D) = (a_1 + a_2)(a_1 + a_2)(a_1 + a_2)(a_1 + a_2) \times (a_1 + a_2)(a_1 + a_2) = a_1 + a_2.$$

The final version of the decision table in the subspace $\{a_1,a_2\}$ has the following form (Table 19).

2.3.1. Decision rules

The above described decision table can also be regarded as a set of decision (classification) rules of the form:

$$\mathrm{a}_{\mathrm{k}_i} \Rightarrow d_j$$

where $\mathrm{a}_{\mathrm{k}_i}$ means that 'attribute ak has value i' and the symbol ' $\Rightarrow$ ' denotes propositional implication. In the decision rule $\theta \Rightarrow \Phi$, formulas θ and Φ are called condition and decision, respectively. Minimization of a set of attributes and values of attributes with respect to another set of attributes simply means a reduction of unnecessary conditions in the decision

Table 16

U	a_1	a_2	d
x_1	2	1	1
x_2	3	2	2
x_3	2	1	1
x_4	2	2	2
x_5	1	1	3
x_6	1	1	3
x_7	3	2	2
x_8	1	1	3
x_9	2	1	1
x_{10}	3	2	2

Table 17

U	a_2	a_3	d
x_1	1	3	1
x_2	2	1	2
x_3	1	3	1
x_4	2	3	2
x_5	1	4	3
x_6	1	2	3
x_7	2	1	2
x_8	1	4	3
x_9	1	3	1
x_{10}	2	1	2

Table 18

	1	2	3	4	5	6	7	8	9	10
1	–	a_1,a_2	–	a_2	a_1	a_1	a_1,a_2	a_1	–	a_1,a_2
2	a_1,a_2	–	a_1,a_2	–	a_1,a_2	a_1,a_2	–	a_1,a_2	a_1,a_2	–
3	–	a_1,a_2	–	a_2	a_1	a_1	a_1,a_2	a_1	–	a_1,a_2
4	a_2	–	a_2	–	a_1,a_2	a_1,a_2	–	a_1,a_2	a_2	–
5	a_1	a_1,a_2	a_1	a_1,a_2	–	–	a_1,a_2	–	a_1	a_1,a_2
6	a_1	a_1,a_2	a_1	a_1,a_2	–	–	a_1,a_2	–	a_1	a_1,a_2
7	a_1,a_2	–	a_1,a_2	–	a_1,a_2	a_1,a_2	–	a_1,a_2	a_1,a_2	–
8	a_1	a_1,a_2	a_1	a_1,a_2	–	–	a_1,a_2	–	a_1	a_1,a_2
9	–	a_1,a_2	–	a_2	a_1	a_1	a_1,a_2	a_1	–	a_1,a_2
10	a_1,a_2	–	a_1,a_2	–	a_1,a_2	a_1,a_2	–	a_1,a_2	a_1,a_2	–

rules, which is also known as the generation of decision rules from the data.

Example 11

Table 19 can be regarded as a set of the following decision rules:

$a_{1_2}a_{2_1} \Rightarrow d_1$

$a_{2_2} \Rightarrow d_2$

$a_{1_1} \Rightarrow d_3$

where d describes elementary sets in D. For Table 11, values 1, 2 or 3 of decision attribute d denote belongingness to the classes 1, 2 and 3, respectively. For Table 12, values 1, 2 and 3 describe elementary sets in D and one can present the above rules in the decision attributes space $\{d_1,d_2\}$ as:

$a_{1_2}a_{2_1} \Rightarrow d_{1_2}d_{2_3}$

$a_{2_2} \Rightarrow d_{1_3}d_{2_1}$

$a_{1_1} \Rightarrow d_{1_1}d_{2_3}$.

Table 19

U	a_1	a_2	d
x_1	2	1	1
x_2	*	2	2
x_3	2	1	1
x_4	*	2	2
x_5	1	*	3
x_6	1	*	3
x_7	*	2	2
x_8	1	*	3
x_9	2	1	1
x_{10}	*	2	2

* Denotes 'do not care'.

2.3.2. New decisions

Logical rules derived from experimental data may be used to support new decisions. Matching their description to one of logical rules can support classification of new objects. The matching procedure may lead to one of four situations [17]:

(a) the new object matches exactly one of the deterministic logical rules;

(b) the new object matches exactly one of the non-deterministic logical rules;

(c) the new object matches no logical rules;

(d) the new object matches more than one logical rule.

In (a), the prediction is straightforward. In (b), however, the rule is ambiguous. In this case, the Decision Maker (DM) is informed about the number of examples (objects in what chemometricians would call the training set) which support each rule. The number is called a strength. If the strength of one class is greater than the strength of the other classes occurring in the non-deterministic rule, one can conclude that according to this rule, the considered object most likely belongs to the strongest class. In case (d), all matched rules are presented to the DM. If all rules indicate the same decision, there is no ambiguity. Otherwise, the strength of each rule is determined and the DM may treat this case similarly to case (b). Information about the strength of each rule can be obtained by presenting the examples supporting a given rule to the DM.

Situation (c) is the most difficult one. In this case, one can help the DM by presenting him a set of rules 'nearest' to the description of the new object. The notation of 'nearest' involves the use of a distance measure [17]. Motivation for introducing 'nearest'

rules results from a belief that giving the DM more information about the neighbourhood of a classified object can give him at least certain suggestions instead of offering nothing. The new object can be identified as an outlier to existing classes or can be considered as an element of a new class.

2.4. Types of attributes

There are different types of attributes.

Quantitative attributes represent measurable properties of objects. Their values are ordered by definition. Examples: temperature, pH, concentration.

Qualitative attributes are expressed in linguistic terms. They can be divided into two classes: (1) *Ordered qualitative attributes*. The values of these attributes can be ordered along an axis of significance. The order of the linguistic values can be represented by a sequence of increasing or decreasing numbers encoding them. Example: polarity = (low, medium, high) can be coded as low = 1; medium = 2 and high = 3. (2) *Unordered qualitative attributes (nominal attributes)*. The linguistic values of these attributes cannot be ordered; in other words, it is impossible to arrange them along any axis of significance.

Application of RST to qualitative attributes is straightforward. For nominal attributes, RST offers evident advantages when compared with other classifiers. To use this type of attribute as an input to classical classifiers, one has to code it in a special way. Each linguistic value is represented by separate input (variable, feature). This encoding, called one-from-*k*, creates a binary vector the elements of which correspond to new inputs. When an attribute takes a particular value, the corresponding vector element is set to 1, while others are set to 0. This type of coding causes a drastic extension of data dimensionality.

Continuous condition attributes present a problem, as in this case, a discretization is required. Both the number of subranges and their intervals have to be optimized. The number of subranges decides about the number of logical rules considered. The number of rules is not given in advance, but is limited by the general requirement that the learning objects should confirm the rules.

There are two possible approaches to the discretization problem. One can optimize coding taking into account only the similarities of the objects in the attributes' space or one can maximize the predictive properties of the information system in the stage of coding. Although the second approach seems to be more interesting it has its limitations. As pointed out by Ziarko et al. in Ref. [18] a low roughness setting, i.e., many small subranges, will lead to weak rules, i.e., rules not supported by many examples, which may be in contradiction to the expert's experience. When the roughness parameter is set high, generalized rules are produced with many supporting cases which lead to strong rules.

3. An illustrative example of application of the rough set approach

In order to illustrate RST, we have chosen to apply it to a QSAR problem. A small data set taken from the literature [19] is used (see Table 20). The example is chosen to be simple for tutorial reasons but it illustrates most aspects of the RST approach.

3.1. Data

The data in question concern modelling of the energy for unfolding of a protein (tryptophane synthase alpha unit of the bacteriophage T4 lysozome), where 19 coded amino acids (AAs) were each introduced into position 49 [19]. The AAs are described in terms of seven attributes: a_1 = PIE and a_2 = PIF (two measures of the side chain lipophilicity), a_3 = DGR = ΔG of transfer from the protein interior to water, a_4 = SAC = surface area, a_5 = MR = molecular refractivity, a_6 = LAM = the side chain polarity, and a_7 = Vol = molecular volume.

The application starts with an appropriate discretization of the information system by translating the values of the quantitative attributes $\{a_1, a_2, \ldots, a_7\}$ and of the decision attribute $\{d\}$ into qualitative terms (Table 20). The condition attributes are coded into four qualitative terms, such as: very low, low, high and very high, whereas, the decision attribute is coded into three qualitative terms, such as: low, medium and high. The qualitative terms of all attributes are then coded using natural numbers. Table 21 gives the range chosen by us for the coded qualitative attributes.

Table 20
Original information system; condition attributes = $\{a, a_2, a_3, a_4, a_5, a_6, a_7\}$, decision attribute = $\{d\}$

Object	a_1	a_2	a_3	a_4	a_5	a_6	a_7	d
1	0.23	0.31	−0.55	254.2	2.126	−0.02	82.2	8.5
2	−0.48	−0.60	0.51	303.6	2.994	−1.24	112.3	8.2
3	−0.61	−0.77	1.20	287.9	2.994	−1.08	103.7	8.5
4	0.45	1.54	−1.40	282.9	2.933	−0.11	99.1	11.0
5	−0.11	−0.22	0.29	335.0	3.458	−1.19	127.5	6.3
6	−0.51	−0.64	0.76	311.6	3.243	−1.43	120.5	8.8
7	0.00	0.00	0.00	224.9	1.662	0.03	65.0	7.1
8	0.15	0.13	−0.25	337.2	3.856	−1.06	140.6	10.1
9	1.20	1.80	−2.10	322.6	3.350	0.04	131.7	16.8
10	1.28	1.70	−2.00	324.0	3.518	0.12	131.5	15.0
11	−0.77	−0.99	0.78	336.6	2.933	−2.26	144.3	7.9
12	0.90	1.23	−1.60	336.3	3.860	−0.33	132.3	13.3
13	1.56	1.79	−2.60	336.1	4.638	−0.05	155.8	11.2
14	0.38	0.49	−1.50	228.5	2.876	−0.31	106.7	8.2
15	0.00	−0.04	0.09	266.7	2.279	−0.40	88.5	7.4
16	0.17	0.26	−0.58	282.9	2.743	−0.53	105.3	8.8
17	1.85	2.25	−2.70	401.8	5.755	−0.31	185.9	9.9
18	0.89	0.96	−1.70	377.8	4.791	−0.84	162.7	8.8
19	0.71	1.22	−1.60	295.1	3.054	−0.13	115.6	12.0

The coded information system is given in Table 22.

The lower and upper approximations are made for each category of the molecules by the set of all attributes, and the accuracy of each approximation is calculated. The results presented in Table 23 are very satisfactory, since the accuracy of all approximations equals one. Hence, the quality of classification also equals one.

The next step of the rough set analysis is to construct minimal subsets of independent attributes, ensuring the same quality of classification, as with the whole set. There are four such *D*-reducts, as given below:

Set #1 = $\{a_2, a_4\}$
Set #2 = $\{a_2, a_7\}$
Set #3 = $\{a_2, a_5, a_1\}$
Set #4 = $\{a_2, a_5, a_6\}$

Intersection of all the *D*-reducts is the *D*-core of the attributes. In our case, it is composed of the attribute $\{a_2\}$, which means that this attribute is the most significant one for classification and therefore,

Table 21
The definition of ranges for attribute coding

Attributes	Codes			
	1	2	3	4
a_1	< -0.115	$[-0.115, 0.54)$	$[0.54, 1.195)$	> 1.195
a_2	< -0.18	$[-0.18, 0.63)$	$[0.63, 1.44)$	> 1.44
a_3	< -1.725	$[-1.725, -0.75)$	$[-0.75, 0.225)$	> 0.225
a_4	< 269.125	$[269.125, 313.35)$	$[313.35, 357.58)$	> 357.58
a_5	< 2.685	$[2.685, 3.708)$	$[3.708, 4.732)$	> 4.732
a_6	< -1.665	$[-1.665, -1.07)$	$[-1.07, -0.475)$	> -0.475
a_7	< 95.22	$[95.22, 125.45)$	$[155.68, 185.90)$	> 185.90
d	< 9.8	$[9.8, 13.3)$	$[16.8, 13.3)$	

Table 22
Coded information system

Object	a_1	a_2	a_3	a_4	a_5	a_6	a_7	d
1	2	2	3	1	1	4	1	1
2	1	1	4	2	2	2	2	1
3	1	1	4	2	2	2	2	1
4	2	4	2	2	2	4	2	2
5	2	1	4	3	2	2	3	1
6	1	1	4	2	2	2	2	1
7	2	2	3	1	1	4	1	1
8	2	2	3	3	3	3	3	2
9	4	4	1	3	2	4	3	3
10	4	4	1	3	2	4	3	3
11	1	1	4	3	2	1	3	1
12	3	3	2	3	3	4	3	3
13	4	4	1	4	3	4	4	2
14	2	2	2	2	2	4	2	1
15	2	2	3	1	1	4	1	1
16	2	2	3	2	2	3	2	1
17	4	4	1	4	4	4	4	2
18	3	3	2	4	4	3	4	1
19	3	3	2	2	2	4	2	2

it cannot be eliminated from the set of attributes without decreasing the approximation quality of classification. The cardinality of a reduct is either 2 or 3, so that in any case five or four attributes are redundant and can be eliminated with no consequence for the classification. The best, i.e., the smallest reducts are sets #1 and #2. Thus, the initial coded information system can be reduced from seven to two attributes. Further reduction can be achieved based on the relative *D*-core and *D*-reduct for the attribute values. The reduced information system can be viewed as a decision table and the set of the sorting rules for the reduct #1 is presented in Table 24.

As set #1 gives the quality of classification equal to one, all the rules are deterministic, i.e., they describe the energy in a unique manner, when certain

Table 23
Accuracy of approximation of each class by seven attributes $\{a_1, a_2, \ldots, a_7\}$

Class number	Number of objects	Lower approximation	Upper approximation	Accuracy
1	11	11	11	1
2	5	5	5	1
3	3	3	3	1

Table 24
Set of sorting rules

Rule	a_2	a_4	d
1	*	1	1
2	1	*	1
3	4	2	2
4	2	3	2
5	4	3	3
6	4	4	2
7	3	3	3
8	2	2	1
9	3	4	1
10	3	2	2

conditions are satisfied. The logical rules found are as follows.

(1) Protein has a low energy of unfolding, if one of the following sets of conditions is met:

-if PIF is very low or SAC is very low

-if PIF is low and SAC is low

-if PIF is high and SAC is very high.

(2) Protein has a medium energy of unfolding, if one of the following conditions is met:

-if PIF is very high and SAC is low or very high

-if PIF is low and SAC is high

-if PIF is high and SAC is low.

(3) Protein has a high energy of unfolding, if the following condition is met:

-if PIF is high or very high and SAC is high.

In Table 25, the accuracy of approximation is presented for each class, when removing one attribute from the reduct #1.

It can easily be seen from this table, how the attributes influence the accuracy of class description. For example, the removal of the attribute a_2 (or a_4) from the considered reduct causes dramatic changes in the accuracy of class 1, and classes 2 and 3 become undefinable.

Table 25
Accuracy of class description when removing one attribute from the system

Class number	RA		
	None	a_2	a_4
1	1.000	0.158	0.357
2	1.000	0.000	0.000
3	1.000	0.000	0.000

Summing up, when applying the rough set approach to QSAR problem, we can find the minimal number of the variables necessary for prediction of the energy of unfolding, coded into tree categories: low, medium and high. The final results can be presented as a set of logical rules.

4. Discussion

RST is a methodology which has demonstrated its usefulness in the context of various cognitive science processes. In particular, RST has provided an array of tools which turned out to be especially adequate for conceptualization, organization, classification and analysis of the various types of data, especially, when dealing with inexact, uncertain or vague knowledge and when discovering hidden patterns and regularities in applications related to information systems and Artificial Intelligence. RST allows:

(a) evaluation of importance of particular attributes and elimination of redundant attributes from the decision table,

(b) construction of a minimal subset of independent attributes ensuring the same quality of classification as the whole set, i.e., reducts of the set of attributes,

(c) intersection of these reducts giving a core of attributes, which cannot be eliminated without disturbing the ability of approximating the classification, and

(d) generation of logical rules from the reduced decision table.

Chemical data sets have not thoroughly been studied in this way yet. However, the above enumeration of the results that can be achieved with RST shows that applications must be possible. It is to be expected that the main difficulty would be appropriate selection of subranges for quantitative variables.

References

[1] L.A. Zadeh, Fuzzy sets, Inf. Control 8 (1965) 338–353.
[2] J.R. Mansfield, M.G. Sowa, G.B. Scarth, R.L. Somorjai, H.H. Mantsch, Analysis of spectroscopic imaging data by fuzzy C-means clustering, Anal. Chem. 69 (1997) 3370–3374.
[3] O.N. Jensen, P. Mortensen, O. Vorm, M. Mann, Automation of matrix-assisted laser desorption/ionization mass spectrometry using fuzzy logic feedback control, Anal. Chem. 69 (1997) 1706–1714.
[4] M. Otto, Fuzzy theory. A promising tool for computerized chemistry, Anal. Chim. Acta 235 (1990) 169–175.
[5] J.M. Mendel, Fuzzy logic systems for engineering: a tutorial, Proc. IEEE 3 (3) (1995) 345–377.
[6] G.N. Chen, Assessment of environmental water with fuzzy clustering, analysis of fuzzy recognition, Anal. Chim. Acta 271 (1992) 115.
[7] G.J. Postma, F.M. Hack, P.A.A. Janssen, L.M.C. Buydens, G. Kateman, A data based approach on analytical methods applying fuzzy logic in the search strategy and flow charts for the representation of the retrieved analytical procedures, Chemometr. Intell. Lab. Syst. 25 (1994) 285–295.
[8] M. Otto, H. Bandemer, A fuzzy method for component identification and mixture evaluation in the ultraviolet spectral range, Anal. Chim. Acta 191 (1986) 193–204.
[9] Y. Hu, J. Smeyers-Verbeke, D.L. Massart, An algorithm for fuzzy linear calibration, Chemometr. Intell. Lab. Syst. 8 (1990) 143–155.
[10] Z. Pawlak, Rough sets, Int. J. Inf. Comput. Sci. 11 (1982) 341–356.
[11] Z. Pawlak, Rough Sets. Theoretical Aspects of Reasoning about Data, Kluwer Academic Publisher, Dordrecht, Netherlands, 1991.
[12] W.P. Ziarko (Ed.), Rough Sets, Fuzzy Sets and Knowledge Discovery, Springer, New York, 1994.
[13] T.W. Collette, A.J. Szladow, Use of rough sets and spectral data for bulding predictive models of reaction rate constants, Appl. Spectrosc. 48 (1994) 1379–1386.
[14] B. Walczak, D.L. Massart, Multiple outliers revisited, Chemometr. Intell. Lab. Syst. 41 (1998) 1–15.
[15] R. Slowinski (Ed.), Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory, Kluwer Academic Publishers, Dordrecht, 1992.
[16] J.R. Quinlan, J. Ross, in: R.S. Michalski, et al. (Eds.), Machine Learning: An Artificial Intelligence Approach, Tioga, Palo Alto, 1983.
[17] R. Slowinski, J. Stefanowski, Rough Classification with Valued Closeness Relation, in: E. Diday, Y. Lechevallier, M. Schader, P. Bertrand, B. Burtschy (Eds.), New Approaches in Classification and Data Analysis, Springer, Berlin, 1994.
[18] W. Ziarko, R. Golan, D. Edwards, An Application of Datalogic/R Knowledge Discovery Tool to Identify Strong Predictive Rules in Stoc Market Data, AAAI-93 Workshop on Knowledge Discovery in Databases, Washington, DC, 1993.
[19] N.E. El Tayar, R.-S. Tsai, P.-A. Carrupt, B. Testa, J. Chem. Soc. Perkin Trans. 2 (1992) 79–84.